DOCUMENT RESUME

ED 076 119 HE 004 030

AUTHOR	Scriven, Michael
TITLE	Evaluating Higher Education in California.
INSTITUTION	California State Legislature, Sacramento. Joint Committee on the Master Plan for Higher Education.
PUB DATE	Apr 73
NOTE	114p.
EDRS PRICE	MF-$0.65 HC-$6.58
DESCRIPTORS	*Evaluation; Evaluation Needs; *Evaluation Techniques; *Higher Education; *Master Plans; *Statewide Planning
IDENTIFIERS	*California

ABSTRACT

There are already a great many bodies devoted wholly or partly to critical scrutiny of California higher education. But it is easy to prove that their combined effort provides only the most superficial evaluation, answering almost none of the basic questions that concern legislators, taxpayers, prospective students, parents and future employers. This report first discusses the grave deficiencies of the present situation, and then reviews the remedial options available. It includes a look at the evaluation situation in higher education and some reference to relevant literature. The point of view of this paper is that good evaluation procedures are necessary for diagnosing the state of California higher education and can, when installed, provide an immense boost towards improvement. There is stress on the beneficial interaction between the evaluation system and the system being evaluated. (Author/HS)

ED 076119

EVALUATING HIGHER EDUCATION IN CALIFORNIA

MICHAEL SCRIVEN

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

HE 00 140 34

PREPARED FOR
JOINT COMMITTEE ON THE MASTER PLAN FOR HIGHER EDUCATION
CALIFORNIA LEGISLATURE

FILMED FROM BEST AVAILABLE COPY

ED 076119

EVALUATING HIGHER EDUCATION IN CALIFORNIA

Michael Scriven

Prepared for

JOINT COMMITTEE ON THE MASTER PLAN
FOR HIGHER EDUCATION

California Legislature
Assembly Post Office Box 83
State Capitol
Sacramento, California 95814

Assemblymen	Senators
John Vasconcellos (Chairman)	Howard Way (Vice Chairman)
Willie Brown	Alfred Alquist
Jerry Lewis	Dennis Carpenter
Ken Meade	Mervyn Dymally
	Albert Rodda

Patrick M. Callan, Consultant
Daniel Friedlander, Assistant Consultant
Sue Powell, Assistant Consultant
Ronna Safonov, Secretary

April, 1973

This is one of a series of policy alternative papers commissioned by the California Legislature's Joint Committee on the Master Plan for Higher Education.

The primary purpose of these papers is to give legislators an overview of a given policy area. Most of the papers are directed toward synthesis and analysis of existing information and perspectives rather than the gathering of new data. The authors were asked to raise and explore prominent issues and to suggest policies available to the Legislature in dealing with those issues.

The Joint Committee has not restricted its consultants to discussions and recommendations in those areas which fall exclusively within the scope of legislative responsibility. The authors were encouraged to direct comments to individual institutions, segmental offices, state agencies - or wherever seemed appropriate. It is hoped that these papers will stimulate public, segmental and institutional discussion of the critical issues in postsecondary education.

TABLE OF CONTENTS

		Page
1.	Overview	1
2.	Strengths and Deficiencies of the Present System	3
3.	Research Versus Evaluation	11
4.	Details and Costs of Good Evaluation	14
5.	Contrasts with Present and Proposed Approaches	20
6.	The Political Dimension of Evaluation	27
7.	Evaluation of Teaching Elsewhere	37
8.	The Quality of Educational Content	39
9.	Personnel Quality Control as a Means to Content Quality Control	42
10.	Direct Approaches to Curriculum Quality	53
11.	Quality of the Delivery System	58
12.	Evaluation of Research	70
13.	Prerequisites for an Evaluation System	76
14.	Nature and Costs of a Minimal Supplementary Evaluation System for California Higher Education	80
15.	Appendix A - Methodology	100
	Appendix B - Comments by Robert Hind and Clark Kerr	107

OVERVIEW

Evaluation is simply the systematic objective assessment of worth, merit or value. There are already a great many bodies devoted wholly or partly to critical scrutiny of California higher education. But it is easy to prove that their combined effort provides only the most superficial evaluation, answering almost none of the basic questions that concern legislators, taxpayers, prospective students, parents and future employers. This report first discusses the grave deficiencies of the present situation, and then reviews the remedial options available. It includes a look at the evaluation situation in higher education elsewhere and some reference to relevant literature. The procedure used is first to "walk through" the requirements for evaluation of one of the crucial aspects of higher education, and then to expand to a more complete coverage by taking up the other dimensions one by one, concluding with a study of some system considerations.

The point of view of this paper is that good evaluation procedures are necessary for diagnosing the state of California higher education and can

when installed, provide an immense boost towards improvement. There will be stress on the beneficial interaction between the evaluation system and the system being evaluated. The annual audit of a business is not merely a record of its welfare at one point in time but also a tremendous stimulus towards avoiding wastage and abuse and towards focussing effort where it will pay off. In short, evaluation is a welfare-increasing intervention, not just a welfare-recording one.

It must be stressed that this paper is not primarily a literature review; it is primarily an analysis and recommendations by a specialist in certain aspects of good management, namely evaluation. Specialists in management, medicine and the law typically provide services rather than surveys though in servicing a commission or committee that service should include enough reasons to make a case for their position.

STRENGTHS AND DEFICIENCIES OF THE PRESENT SYSTEM

Evaluation is not research-oriented, it is service-oriented. An evaluation system for California higher education should be able to service the crucial decisions and questions of the governing and 'consumer' groups involved. For example, it should be able to answer such questions as the following (the evaluative terms are underscored):

> How does the quality of higher education compare with other states?
>
> How does the teaching (of a given subject at a given level) on one campus of a multi-campus system compare with that on another (in cost-effectiveness terms if possible)?
>
> How valuable is the curriculum covered in a given professional school or department for students who go on in that profession?
>
> Does the use of community colleges as feeders for the four-year systems harm the transferees significantly in competing with those going to the same campus for four years?
>
> Is the quality of undergraduate teaching suffering significantly on those campuses which have heavy commitments to graduate education and/or research?
>
> Is there adequate identification and reward for good teaching or administration or public service?
>
> Is adequate service being provided to minority groups, working adults, women and the poor?

> Are the faculty/administration (or particular faculty/administration in the different components of California higher education) being overpaid or overworked; by what standards? Are their working conditions or those of students reasonable or exploitative?

The kind of value judgments involved obviously leave room for some element of subjectivity. But that is a poor excuse for not trying to reduce the subjectivity. To take a simple example from the performance-contracting literature on elementary education: we have good reasons to support the claim that the best of the independent contractors is from three to ten times as cost-effective as traditional remedial reading programs. Now it's true that studies supporting this conclusion use certain reading tests and not others, and there is room for some argument about whether those tests are ideal. But they are chosen by the schools, with parental involvement, and it's pretty reasonable to say they're as good as any. And there's no reason to think results would be different on any other tests. So assessments of quality, cost-effectiveness, etc. are by no means just subjective. And a good deal of higher education is concerned with teaching subjects that

are perfectly amenable to exactly the same analysis as remedial reading (e.g., math, science, foreign languages). It's just that such studies have never been done.

Not one of the above listed questions can be given an answer of even moderate objectivity at this time.

It is not easy to provide the kind of evaluation system that will answer these questions with reasonable objectivity. It is not mere negligence that has kept it from being installed. Such a system may require some resources to initiate. Still, it can be expected - indeed required - to pay for itself many times over. Its justification lies not only in the efficiency and economy it inspires, but also in the fact that it demonstrates accountability. There will be occasions when the state or the citizenry will feel the need for an appraisal of the worth of some program or component, whether or not there will be a net savings from the evaluation. As the saying goes in another context, it is not enough that justice is done, it is necessary that it be seen that it is done.

It would be fair to say that the analytical and coordinating agencies we have at present (e.g. Coordinating Council for Higher Education, Legislative Analyst, legislative committees, Department of Finance) do a pretty good job of picking up on unsound fiscal management procedures and on certain gross kinds of redundancy. Their prime deficiency consists in their almost complete absence of quality output controls, not in lack of fiduciary, i.e., input controls. Without a good quality-control system it is impossible to identify the significant effects - if any - of cuts and increases in support. Hence, one cannot establish the possibility of maintaining quality at reduced cost, or increasing quality for minimal cost. Without quality controls, political and fiscal considerations naturally rule the field with unknown effects on quality, each side claiming triumph (or disaster) and dismissing the claims of the other as mere prejudice.

What would serious quality control mean for California higher education?

There is one respect in which quality is maintained by a very good - though not flawless - system. That is the University of California's

system of selecting personnel for high intellectual performance in their specialty - the so called "research emphasis." This procedure, as used in main-line departments, with some exceptions and reservations (discussed later), is extremely careful, and has kept UC very high in national rankings which tend to emphasize exactly this dimension. But the smartest researcher in the world is obviously not thereby a good teacher, indeed there is a well-advertised risk that he will be less of a good teacher. And the procedure that works for UC (and the big private universities - USC and Stanford) has worked less and less well as one looks at the California State University and Colleges and the California Community Colleges. Moreover, it is an extremely expensive system; it is slow to operate (and hence costs us many appointments), and it requires vast time expenditure by high-level personnel (further details later). Let us now focus on an area of even greater social and educational significance than the selection of highly-trained staff.

The basic mission of higher education is to

teach students something worthwhile. We will spend some time exploring one simple example of what it would take to see if it does this. This will show just how far we are from having even a basic evaluation system.

Consider a subject where there is little room for dispute about the value of the content, e.g., freshman math for science majors. Now it's pretty easy to prove that everyone in that class learns something useful from the course. This fact is probably the one that contributes most to giving the instructor the nice warm feeling that he's doing something worthwhile. To the eye of the professional evaluator, however, that fact is totally uninteresting. For the evaluator knows very well that the same fact would hold if those students had gathered in that classroom on the same schedule - but without an instructor. Not only do students learn from textbooks, but they can read far more than they can hear in a given period, and they can read it at whatever pace and with as many repeats as they need. So the unaided text is often a formidable competitor for the instructor. Even more formidable is the

combination of two texts, written from different viewpoints, and one low-priced teaching assistant (TA) to answer individual questions and maintain order. If one or both of those texts is properly developed and programmed, the instructor's chances of "winning" by lectures almost vanish.

There are a number of situations in which the instructor is indispensable - or valuable - for example, in presenting a graduate survey course on recent research literature, or in inspiring students with a vision or an approach that has more life and juice than the written page. How many instructors can justify their salary in these terms? We do not know. Faculty of course have other duties connected with instruction - such as grading papers, selecting scholarship students, job and study counseling, etc. But are they cost-effective in the basic role of classroom teaching? No one knows - nor has anyone in California higher education made a serious effort to find out.

There is no significant scientific difficulty about answering the question of the instructor's classroom utility by comparison with the "text

plus TAs option." In fact, Keller's work in Arizona gives us a very strong hint that even good instructors will be easily beaten.* Are there significant exceptions? Will this approach work on a "steady-diet" basis as well as a "one-shot" basis? That we don't know. But the evaluation system for a multi-million dollar educational instrument certainly ought to tell us.

There are some cost difficulties and substantial political difficulties with getting these answers. These difficulties will be analyzed below. But first it is essential to discuss a more general possible difficulty. The answer to this has considerable effects on the cost and political problem.

---

* See, for example, Fred Keller "Goodbye Teacher. . ." in Journal of Applied Behavior Analysis, Spring 1968.

RESEARCH VERSUS EVALUATION

Reading the above account of the kind of answers an evaluation system should be providing, one might think it sounds like a recommendation to do a great deal of basic research on teaching procedures. It is true that if we had the results of that research, the evaluation could be done with almost no expenditure. It is also true that the failure of the universities, which have done research on almost everything else under the sun, to study their own basic procedures with the same care shows an outstanding lack of responsibility. However, evaluation does not require the research results. Evaluation can be of particular specific performances - it does not necessitate any degree of generalization at all. We can evaluate the performance of a particular drug on a particular patient, or the use of a soil additive on a particular field, without coming up with a single generalization.

Similarly, we can compare the performance of the teacher thought to be the best in his department by his peers and students on a particular

campus against (a) the base texts the department has chosen to use assisted by an "average" TA (b) the instructor's image on TV replay (c) the TA alone. We will not have any hope of unearthing a grand generalization, only a very significant fact. For if the individual thought to be the best teacher is out-performed by the text plus TAs (to take up just the first of these comparisons), the very real possibility is created that the other instructors are still worse. We can certainly say that either they are worse or their method of identifying the best teacher is faulty. We can't say whether this would be so with any other text, with all other subjects, or with the student body on every other campus. But we have an excellent basis for a rather jaundiced evaluation of the way this department is teaching this subject.

It would seem appropriate to call this applied research, but it may lack even the generalizability of applied research in engineering and the sciences - where the results of testing one sample of insecticide or an antibiotic or a photographic emulsion can be generalized with considerable confidence to other such samples. On the whole, it is better

to talk of "evaluation research" or just "evaluation" rather than using the pre-existing and somewhat misleading term "applied research."

DETAILS AND COSTS OF GOOD EVALUATION

Continuing with our simple little example of evaluating the merit of the usual approach to undergraduate instruction in California higher education, we may ask whether the cost would be prohibitive. The answer is no. If we tried for a large control-group study with a long-term follow-up, it might be. At this stage, however, we are principally interested in very crude results. If we do not get a large short-run gain from the use of very expensive teaching personnel, we should not be using them until someone finds another reason for doing so. (Of course, some of them may have to be retained for other purposes or classes). In doing evaluation, we do not need an external control group at all. We only need an instructor who is interested in improving his teaching. It is easy to give him full instructions so that he can actually do the evaluation himself, at no extra cost at all. For credibility and reliability purposes, we would do better to assign him a few hours of professional assistance, in the form of someone who would administer the

tests (set by the instructor) - and that is all. (We return to a quantitative estimate of costs in a later section.) What would be involved? Leaving out practical complications, we can describe the three key elements very simply. (A somewhat expanded description of this kind of evaluation investigation can be provided on request.)

1. A pretest, taken on the first day of class, covering the material the course will cover, at exactly the same level of difficulty as the final exam. Without this, one has no idea whether some students are coming in with enough knowledge to get a "C" without any contribution from the teacher or textbook. (The results of the pretest are not seen by the instructor, not graded until the final is done, and not graded by the instructor.

2. Division of the class into two halves (e.g., by numbering the class list and putting every odd-numbered student in the control group.) One half meets to read the text plus supplementary material, with a TA on hand to check attendance and answer individual questions. The other half gets the usual lectures. Attendance is also taken to prevent "leakage" into the latter group from the control group.

3. A final exam, matching the pretest in difficulty. This is easily done by having the instructor compose, or pull from his files, several exams before the course begins, and having the evaluator toss a coin to decide

which is to be the pretest and which the final. It is preferable not to let the instructor know which was used for the pretest, so as to avoid (unconscious or conscious) "teaching to the test", i.e., avoiding coverage of pretest items so as to focus on the others. The final and pretest are graded simultaneously by any competent person selected by the instructor, as long as it isn't someone who has taken or studied the particular version of the course given by this instructor. (In return for this favor, the instructor undertakes to do the same for one of his/her colleague's courses.)

Given this data on the two groups, we can readily determine whether the lecture method is worthwhile in this case. The experiment as it stands will sometimes be rather heavily biassed in favor of the instructor, and the evaluator needs to check on this possibility in any case where he/she will use this design. It will work well where there is a departmental choice of text and topics, and the exams are fairly standardized to cover these. It will not work well if the instructor covers a great deal of material in lectures and exams that is not in the text. There are, however, many situations where these conditions are met, besides departmental introductory

courses (for example, where the instructor uses his/her own text, or issues class notes or a very comprehensive reading list, or where an organization like Fybates publishes professionally-taken notes from the course). So this design will work well for many of the most important courses in California higher education.

Where it will not work at first, it can be made to work quite soon as follows. (And this is one of the ways in which evaluation procedures improve teaching.) Given that the instructor is chosen as one who wishes to improve teaching performance, encouragement is provided to write adequate class notes, with the aim of covering the material not covered in the published readings to the required level of understanding. This psychological basis for encouragement is the possibility of a new and more attractive as well as more effective teaching role as advisor, resource person, and individual or small group counselor. Such a role which will open up if the large-lecture format can be replaced by reading. (It should be stressed that these procedures, strange though they sound to most faculty and students are not

speculative but tried and true, not only by Keller but by many others for many successful years.)

It will be clear that we are not merely suggesting a possible critical appraisal of present teaching procedures, but looking towards improvements. For the basic handicap of the large lecture course, with a wide level of ability and background among the students, is that it can only proceed at a single rate. It cannot be backed up for a repeat for the student who missed something crucial; it cannot be speeded up for the student who finds it repetitive or knows part of it already. It is in the individualization of pace that the move towards written materials has the advantage. The instructor then has the option of delegating supervision to TAs (with him or herself preparing or updating other course material), continuing in the individual tutorial role, or running small discussions or tests for those who are at the same point in the notes. Audio-visual transcripts of lectures provide the same advantage, but at higher costs and (to date) no proven gains. However, the Legislative Analyst was quite correct in suggesting

that there would probably be a gain in cost-effectiveness from the use of CCTV for lectures, since the relatively small loss that has appeared in the few studies done, and the (mostly) one-time hardware cost, is more than compensated for by salary savings over a few years.

Precisely analogous investigations can be undertaken to handle the other evaluative questions in the original list, and other dimensions of the present question. For example, it's obvious how to use this design to cover the comparative efficacy of several teachers, or texts, or procedures for selecting students. The more general questions about overall merit simply require a more extensive base of particular evaluation activities like that discussed. A systematic approach will be outlined later.

CONTRASTS WITH PRESENT AND PROPOSED APPROACHES

It is extremely important to see the vast differences between the kind of evaluation just described and what goes on at the moment.

A few years ago, as is well known, there was very little effort at any level to evaluate teaching. What effort there was distinguished the California Community Colleges and the California State University and Colleges rather than the University of California. But even there, the emphasis on teaching is superficial and more a matter of there being less emphasis on research than a real case of evaluating teaching. In recent years, partly from public pressure, there has been a substantial change. But the evaluation involved is still absolutely trivial. A fair description of the procedure used now in the typical main-line department of the UC system, after several years of heavy pressure from the administration, still has three fatal defects and shows no sign at all of ever transcending them.

1. It is heavily influenced by the judgment of colleagues who have never visited the classes of the instructor

in question, or - if they have visited - have not visited a reasonable sample, say 10%, of the lectures in any course. Where no class has been visited, the judgment may be based on hearsay from colleagues, on gossip from a very atypical group of students (usually a few gossipy majors in that department), on the instructor's performance in presenting a paper to a departmental coloquium or at a professional meeting, or simply on a judgment of his personality or research competence.

2. Even if the above limitations on the evidence used by the judges were removed, the judgment would still be made by other faculty members whose capacity either for judging or doing good teaching is completely unknown. It is very nearly a case of the blind leading the blind; at least we have no way to tell that it isn't such a case.

3. It is natural enough to handle difficulty (2) by placing more emphasis on the judgment of the ultimate consumer of instruction, the student. But the procedures for doing this, quite apart from the problem of sampling and inexpert interpreters, are essentially unsound for other reasons. The evaluation forms typically used call for some judgments that the student is not in a good position to make, e.g., judgments on the professional competence of the instructor, and for this reason are often scorned by the faculty. But a more fundamental problem is that the qualities the student can judge reliably (e.g., friendliness, willingness to accept criticism, setting heavy assignments) are not related in any known way to effective

> teaching. Of course, many people are quite sure that lots of assigned reading is good, that being friendly or accessible to students is good, etc. There is absolutely no evidence for this, and there's a heavy cost involved, in student and faculty time respectively. Heavy reading means less time to concentrate on core reading and study class notes. "Open Door" offices are the ones where the instructor has less time to prepare complete notes for the next class meeting.

The principal basis for evaluation of teaching must be measurement of the teacher's actual contribution to student learning. We do not have to have a fancy theory of human training to find that out, only some good pre- and post-tests.

So much for present procedures for evaluating teaching in California higher education. Although no such studies as suggested in this paper are done - unless as secret research - the University of California (for example) persists in making claims like the following:

> The University serves as a model or laboratory in which new modes of undergraduate education may be developed and tested.*

Who ever heard of a laboratory in which the product or the process isn't objectively tested? The University apparently feels that the standards of objectivity of its research labs should not even

* UC Growth Plan 1972-82, approved by the Regents at the July 1972 Meeting.

be attempted with respect to its own performance in classrooms. Recently it has been suggested that we should move towards a different system for evaluating teaching, or faculty performance in general, namely classroom contact hours. The state university of Texas moved to such a basis a year ago and we have some preliminary evidence about its success. But that system involves very complicated offsets from a rather high basic load. The system favored in California involves a lower but rigid load - but one that is about 50% higher than at present.

Such an instrument has certain attractions as a basis for evaluation, notably its simplicity and the fact it might involved as much as a 33% reduction in faculty size. Every two instructors pick up a full extra load between them, thereby eliminating one out of three positions. But, of course, this completely ignores quality. It isn't any good assuming that there will be a large, let alone a 33% drop in quality (or quantity of students taught or material covered). But it would be equally absurd to assume there would be none, or a negligible amount.

That such a system could be considered in the absence of procedures to check quality loss is an unfortunate sign of the general lack of willingness to treat higher education as a fit subject for serious study. The colleges won't do it because they do not really value their teaching function as such and/or because funds for such studies are relatively limited. The state won't do it because it sees a substantial chance for a massive cost-saving without doing it. This brings up the topic of the next section - the politics of evaluating California higher education.

In concluding this section, however, it may be worth listing some of the defects with the straight classroom contact-hours approach.

(a) Class hours make lab supervision, which usually requires little or no preparation, equal to lecture time - or omits it, which errs in the other direction.

(b) Class hours does not cover the number of office hours; adding them in would not work since one could list 25 a week and only be seeing two students.

(c) Class hours does not cover preparation time. Some instructors keep their courses updated by discarding almost all previous

materials and repreparing each time. Preparing introductory French language or freshman calculus may not involve such work, whereas fast-changing subjects, e.g., Asian politics, recent issues in constitutional law, computer design, may necessitate it.

(d) Class hours makes no reference to class size. Hence there are no advantages offered for admitting 300 instead of 30 or 100, which throws away the chance of massive savings. Where such allowance <u>is</u> made (as in Texas) it doesn't distinguish the instructor who lectures just the same to 300 as to 30 and leaves the rest up to the TAs. In fact, since the TAs are often told to set - and nearly always told to mark - all assignments, the big class can be easier. For the instructor who meets with his TAs regularly, randomly visits section meetings, directs all grading, prepares special materials or readings for the groups that emerge from these classes with special difficulties, etc., an allowance beyond contact hours should be made. The University of California makes a point of its special opportunity within California higher education to provide "supervised apprentice teaching." But it provides absolutely no incentive nor even the requirement that its TAs <u>be</u> supervised. The commitment is empty.

(e) Class hours makes no allowance for courses that are repeated during the same day or week, the usual device for reducing real load in places with unrealistically high contact-hours requirements, e.g., the State

University and Colleges and the community colleges.

(f) Class hours makes no reference to the extent of class involvement - being chairman at a rap session requires little work and may produce little learning, whereas an hour's lecture on the significance of revenue sharing may take a week of steady reading to prepare. Conversely, the discussion may be far more valuable in certain cases. This leads to the crucial weakness of contact hours.

(g) Class hours takes no account of either student reactions or student learning gains, so it is necessarily a poor basis for evaluating an educational system, as well as a poor basis for evaluating instructor workload. It is just another input measure, like the percent of Ph.Ds on the faculty.

THE POLITICAL DIMENSION OF EVALUATION

In the early days of time and motion studies, operations research, incentive pay, piece work, ergonomics, and socialist planning there was an overemphasis on the "new science," whatever it happened to be, at the expense of existing humanity. In short, these well-intentioned reforms were often exploitative. The unions have traditionally provided the counterpoint to the efficiency experts in this long dialectic, and from it we can at least learn some of the mistakes to avoid.

Take the proposed use of the contact-hours criterion as an example. The political appeal of imposing, say, a nine contact-hour requirement on U.C. faculty, or twelve hours on CSUC faculty, depends entirely on the idea that contact hours are a pretty fair measure of working hours. Working hours are the tangible index of performance for which the salaried accountant or typist is paid. Beyond this, in large offices and bureaucracies, only a minimal level of quality is required. It is understandable to the citizen

that the school teacher should do rather less than forty hours in class, because he/she is at the school for a good many hours more than that, and takes work home on top of that. But to the average citizen, the professor simply vanishes, apart from classes and possibly a few office hours. The public is simply not persuaded that he/she is really working hard the rest of the forty hours. Surveys which reveal what the faculty <u>say</u> they are doing are scarcely persuasive to anyone over the age of ten, whether or not in point of fact they are fairly accurate. A few rumors about outside consulting jobs at princely rates and the use of TAs to avoid the hard work (in the U.C. system) go a long way to convert envy into hostility.

Add to this the explosion of violence on the campuses, plus heavily increased taxes, and the result is predictable - no cost-of-living compensation, let along real raises for faculty.

There is nothing surprising about this situation nor particularly unreasonable about it. State-financed higher education in this country has simply never recognized its obligation to

show the state that it is operating at a reasonable efficiency level, indeed it has scarcely even shown the least interest in exploring the question for its own sake. There <u>are</u> powerful arguments and analogies to support the instructor's defense that he/she is working hard. They have never been made a major part of the college's public relations effort - chiefly because scarcely anyone on the campuses has ever thought of them.

For example, one must face the fact that the college instructor is the interface between the thousand or so technical journals in his professional field (and the twenty to fifty significant books every year) and the student. If higher education is to be passing on to students the present state of knowledge, it must be updating most courses most years. Now if you assigned someone to produce a decent synthesis and translation of several hundred journals and a few dozen books into relatively elementary terms - the result to be organized into a series of thirty lectures - he/she might well feel it wasn't a bad year's work. If we now said, make it a hundred and fifty lectures and cover rather

more reading (e.g. from related fields), sort it out into different dimensions and organizations, set ten or twenty examinations on it, and grade a few hundred assignments - it could look like a pretty good year's work even without committees, writing, interviewing, professional meetings, and curriculum planning, etc.

This kind of rhetoric is part of evaluation; the presentation of an appraisal is often more persuasive than the percentages in it. And the way these matters are presented has an incisive influence on their political feasibility. If you can sell the increase to nine contact hours for U.C. faculty (twelve for CSUC) to the electorate as a modest enough substitute for the forty of the working staff, you have a great plank in your platform. If the electorate thinks that classroom hours makes no more sense as a standard for faculty than sermon hours do for clergymen, or public speaking hours for a state legislator, then you'll drop it. California higher education has a great deal of fence-mending to do on this kind of issue. And friends at court can do

a great deal for the system if they can focus some of the public enthusiasm for reform into quality-controlled reform.

In the present political climate the two principal dangers are that some superficial evaluation device such as the contact hour standard will be imposed, with serious effects on morale and probably on quality; or that it will be successfully opposed (or repealed) as a sweat-shop practice by a political alliance of faculty unions or associations with the rest of the labor movement. Neither outcome promises significant benefit to the student and hence to the state whose welfare is eventually dependent on the medical, social, technical, agricultural, managerial, and humane education of its college graduates. The probably short-run savings of tax dollars from the contact-hour approach would be bought dearly if there is even a five percent decline in the capacity of the average graduate to judge the merits of a complex social reform proposition on the ballots ten years from now, for such a margin is often enough to effect or bar passage. (We shall turn to the relevance of

higher education to such decisions in a moment.) The main point of the present discussion is to stress the fact that whatever steps are taken towards improved or new methods of evaluation, there will have to be accompanying educational campaigns, or they will be doomed by the present very low level of understanding of the nature of higher education and political realities.

Turning from the taxpayer to the faculty, one basic political fact is that people naturally resent the kind of basic criticism, of the way they are doing what they are paid for, that is implicit in this paper. The political reality may be that they are in for a speed-up and the only choice they will be able to effect at all is the way it's done. Or the political reality may be that the emotional connotations of "speed-up" are still strong enough to pull out decisive labor opposition to any such requirement. It is not easy to see which way the ball will bounce in 1973-4. All this report can do is to point out the political implications of different approaches to evaluation, and try to avoid recommendations which would be politically absolutely impossible.

To take one example, faculty are not going to make large-scale moves towards converting their lectures into extremely thorough written materials if those materials will be used to make the faculty redundant. Where it turns out that such a shift is producing very substantial educational gains (Keller's big classes in introductory psychology went from something like 65% getting 65% on the final to 85% getting 85%), it will become a political problem to make the shift - a problem in the micro-politics of the campus. The easiest solution will be to reward the gain in quality by a teaching award without attempting to make a short-term economy out of it. The next easiest would be to offer incentives that share the gain in productivity with those instructors who will handle larger classes in this way. The next would be to go for the curriculum changes that reduce some of the redundancy that is presently necessary in high level courses in order to handle the extremely irregular achievement level of those who have been through the introductory course. This could result in fewer upper division courses (for the

same coverage) or new courses to widen the coverage (for the same number of faculty). A deal would probably have to be made to guarantee that the savings would not come out of the hide of those who now have the jobs. They must receive some of the benefits of the productivity gains they facilitate, even though the system may not replace all of them as they retire. General considerations of social utility may of course favor providing more jobs and a better education rather than tax savings and the same education. That will be a political decision, though it's an economic plus a moral problem.

To take another example, it is clear that the standards of work quality applied to the average clerk typist or casework supervisor are pretty minimal and mostly non-quantitative. It is politically unlikely that one could introduce the very high powered task-related evaluation systems involved in incentive pay or piecework approaches to the state or federal civil service. It might well be argued that there is a far greater difference between the work required by an associate professor teaching introductory genetics for the

tenth time and one teaching contemporary lunar geology than between any two same-level civil servants in different departments.

But it's a fair bet you'll never get pay differentials established on the campuses to compensate for that kind of difference. Still, it might be possible to get them in for proven differences in teaching performance, just as has long been done for differences in research performance. The political trend has been the other way, beginning with the big New York City system where all quality differentials were eradicated in favor of automatic step increases, following the K-12 model. California was not far behind, and with the nominal exception of the UC system, has adopted this "democratic" approach from top to bottom. But there are some faint signs of reversal, beginning again with the K-12 system in those states which are introducing various types of accountability. The practice, sporadic in the past, of systematic merit pay for master teachers, tied to performance, is emerging again in the plans of states

who feel they are now strong enough to buck the union opposition.

It is impossible for an evaluator not to conclude a section on the political dimension of evaluation without pointing out that almost all political influences on evaluation procedures have been the result of deals between teachers, unions, and the so called power holders (boards, administrators, and legislators), and have resulted in heavy costs for the students and hence for the long-run welfare of the society. Quite obviously, systematic evaluation of the teaching mission of California higher education has never been paid more than lip service.

While the situation is not much better in other states, it is worth noting that only the three largest fiscal units in the U.S. have moved to the 'lock-step' system for higher education salaries, excluding virtually all consideration of merit (N.Y. State, N.Y. City, and California). The only defense for this is the claim that no objective basis for discrimination is available. This is not only false, it is a cynical confession since any one of these systems has the resources to go it alone in developing such a basis. The significance of the tally is simply that there has so far been very little political gain to be had from improving the evaluation of teaching. It would be exciting and very worthwhile if California could reverse this trend.

However, the most important comparison in this instance is not with other states but with other countries. There is just one feature of college management abroad, little remarked-on these days, that provides a most important improvement in the evaluation and indeed the per-

formance of teaching. It is standard operating procedure in most universities in, for example, both England and Australia (where all universities are state-supported), that the final examinations in every subject are set and graded by someone who did not teach the course. This immediately opens up the possibility of estimates of comparative efficiency across faculty in achieving the same goals. And it produces the instructor's best efforts, because he/she knows the results are going to be judged by someone else. The procedure has been introduced to this country. Swarthmore College has run its honors program, involving 50% of the college's enrollment, with outside examiners for over thirty years. The judgment of these examiners is heavily used in the decisions on promotion/retention of faculty.

THE QUALITY OF EDUCATIONAL CONTENT

When we began to particularize about the evaluation of teaching, it was stipulated that we should consider a course where there could scarcely be any dispute about the utility of the content. It is now time to remove this restriction and look into this question of quality. The quality of the delivery system, of which the teaching component is perhaps the most important element, is sorely in need of evaluation. But the quality of content is in little better shape. Again, California higher education has been content to drift along the same path as other states. This is the more unfortunate since California has long been acknowledged as a leader in the development of a comprehensive system of higher education.

Any across-the-board attempt to improve quality of content in higher education is up against truly formidable difficulties, considerably more formidable than those that bear on improvement of teaching. For, while the college teacher is prepared to concede that his or her

teaching may stand in need of some improvement, he/she is not about to concede that his/her judgment of material in the field of his/her speciality can be improved. Not only are these difficulties hard to overcome, but there are serious reasons for preserving the autonomy of the teacher - up to a certain point. The method for quality control of content that is normally used in higher education is simply quality control of personnel. Although there is some departmental supervision of course content, and some joint selection of courses, this is by no means the majority procedure. There are places within California structure where public demand quite quickly manifests itself in new curricular items - notably the community colleges, which are rather responsive to job market needs in their community, and U.C. Extension, which has to cover its own costs. But by and large, the instructor, having been selected for quality, is then left on his own. Even if the title of the course is prescribed, the choice of text (usually), and the treatment/approach/examinations are up to the instructor. The professional schools, with their

somewhat more tightly structured curriculum, place correspondingly greater constraints on the freedom of those instructing freshmen, but even there the situation with respect to upper level courses, and certainly graduate courses, is one marked by a great degree of autonomy.

No evaluator could have any basis for arguing that he/she personally, or indirectly, can identify judges of content that are better than the collective California faculty. However, the judgment is not made by this collective faculty. It is made by a few individuals, with rather limited supervision. Not surprisingly, there are some very serious weaknesses, often quite notorious on a given campus. The basic question for the evaluator is whether this is simply part of the cost in inefficiency or lack of quality that one must pay for any system (since each will have some weak points) or whether it is an exceptional, separable cost that could be reduced by switching to another system. There is no evidence about the success of radical alternatives to this system, in this country, on this scale - but that is no grounds for not exploring some alternatives.

PERSONNEL QUALITY CONTROL AS A MEANS TO CONTENT QUALITY CONTROL

Let us begin looking at the present indirect method of insuring quality of content. It has two major loopholes and one small one.

The first of the giant loopholes is the actual failure to apply standards of real quality in the selection of personnel. In my judgment, this occurs in some (but relatively few) departments at U.C., U.S.C. and Stanford, in many cases at the State University and Colleges and on a very large scale in the community college system. (It should be stressed that substantiating these judgments with absolutely water-tight evidence would cost something like the operating budget of De Anza College for a year. But their plausibility can be substantiated with data presently available. For the purposes of this report the most important impact arises from their possibility rather than their probability. The main point of the report is to show the need for an evaluation system that would tell us whether such possibilities

occur.) The weakness in the personnel selection procedure for controlling the quality of curriculum occurs in, e.g., the community colleges in the following way. Suppose that they are in the process of considering a large number of applicants for a particular position, as is the case in the present job market. A number of these applicants have Ph.D.s from very good institutions of higher education. A number of others have the Master's Degree only, and yet others may even lack that. Now the present situation is that the candidates with the Ph.D. are unlikely to get the job.* Of course a reason is given for this, for example, that the Ph.D. people either are not so interested in teaching, or are less likely to be "good investments" because they'll move on as soon as they can, or that they are "too narrow" or "too expensive." Those reasons will be considered in a moment, but note that we are at the moment only considering the question of controlling the quality of content. There can be little doubt that holders of the advanced degree, not through any special virtues of the thesis itself, but because of the arduous selection proce-

* See the figures on candidate qualifications and employment rates in John Huther's detailed study, "Small Market for Ph.D.s - The Public Two Year College", AAUP Bulletin, March 1972, pp. 17-20.

dures and the intensive scrutiny given during at least the one or two occasions of examinations during the Ph.D. process, do militate quite strongly in favor of a higher degree of intellectual rigor and knowledge of research problems, procedures, and findings. It is not being suggested for a moment that this further knowledge is what should be taught in the community college. It is merely that the mind of a man or woman who has been trained to this level is, other things being equal, better equipped to distinguish poor theories and hypotheses, poor scholarship, and poor criticism from their better counterparts. This is the most important respect in which the use of personnel selection can control the quality of content. Unfortunately, one must conclude that a great deal of inverse snobbery goes on in the selection process, or at least that it may. Without an evaluation system, such a possibility looms large. This possibility is that the community college faculty does not like the idea of having around somebody whose intellectual prestige and perhaps whose intellectual attainments far outshine their own, potentially if not

actually. They are afraid that there will be trouble over the imposition of quality standards that they and their students cannot manage. Are their fears really ill-grounded? One can only say that there is no systematic evidence, and what anecdotal evidence there is must be regarded as equivocal. Certainly there is little risk involved in taking on one or two members in a department at this stage when extremely high intellectual quality is available quite easily, since in a minority position they can scarcely do much damage. As far as the students are concerned - on whom they are supposed to be imposing arbitrarily high standards - there is little that serious counselling of the new instructors by the chairman cannot handle. There's little evidence that some increase in the standards will have a net negative impact, and there is the perfectly simple recourse of not renewing the contract in the first two years.

It is impossible not to conclude that there are some fairly serious failures in the use of personnel selection as a means to improve quality of content.

However, there is an important respect in which the fears of the community college faculty in such cases have not yet been done full justice. And this brings us to our second major loophole.

There is an extremely important error in the development of curriculum materials that has become especially evident in the past two decades of intensive new curriculum material development, ignited by Sputnik. It has been referred to as the "academic fallacy." The fallacy consists in supposing that the most logical method of structuring an area of knowledge, such as mathematics or physics, provides the best or at least a satisfactory basis for pedagogical developments. The most obvious example of the academic fallacy is the New Math, which to a large extent involves imposing upon the curriculum the structure of mathematics as seen by the professional mathematician. It wasn't impossible that set theory should have provided a workable approach to curriculum development in mathematics, but it would have been rather a long shot, if one were thinking about the matter purely from the educational point of view. After all, the kind of

conceptual ability that it takes to appreciate set theory may not be present in most students who are anxious to acquire, or need to acquire, basic mathematical skills. Now, in the present state of neglect of teaching, it is indeed the case that the highly intellectual approach to a subject, characteristic of those who have a Ph.D. in it, tends to go along with wholesale commission of the academic fallacy. And community college faculty are quite right to be nervous about the idea that the most abstract approach, impeccable as it may be in a graduate school, may well be a poor vehicle for learning at the community college freshman level. As a matter of fact, it is almost certainly a very poor basis for teaching freshmen even in the high-powered university components. So this second loophole in the process of controlling the quality of the curriculum amounts to saying that "quality" is not merely a matter of respectability in the eyes of the advanced professional. It must also take into account appropriateness for the level of instruction.

Two apparent conflicts between this line of

argument and that presented in connection with discussion of the first loophole should now be taken up. In connection with discussion of the first argument, it was proposed that intellectual quality is one way of controlling the quality of the curriculum. It now appears that this is being denied. The point can be clarified in the following way. It takes considerable intellectual training, perhaps even the more the better, to provide somebody with the capacity to distinguish the tawdry from the sound theory or interpretation. But this ability by no means guarantees freedom from other faults in selecting and presenting materials. In short, higher intellectual accomplishments are a desirable but not a sufficient quality in faculty members.

The second apparent inconsistency is connected with the first. Having criticized the community colleges for failing to pick Ph.D.s, has not the argument then gone on to give the perfect reason why they should refrain from doing so? No so; for there is a simple solution. The community college faculty have an excellent point to make in arguing that they do not wish to see

the academic fallacy visited upon their students. On the other hand they have an obligation to select the best possible personnel. The way to combine these is by a more substantial infringement on the autonomy of new faculty than would be appropriate at a graduate school. That is, by selecting and then supervising the instruction done by "hot shot" new faculty. We cannot avoid the problem of the unreliability of the senior faculty in providing such guidance. But we can reasonably expect them to have a better sense of what is teachable and/or useful than the new Ph.D. At least they can provide criticisms which the new instructor can accept or meet by careful testing of his or her students.

As for the claims about "narrowness" and "cost," these are on a par with racial considerations. People deserve to be treated on their individual merit. Some Ph.D.s are narrow in their interests, but many have Master's degrees in one field and a doctorate in another; one can hardly prefer someone with a simple M.A. over such a candidate on the general grounds of excessive narrowness! There are many Ph.D.s whose entire

interest is in abandoning the confining narrowness of their research and turning to their many more general interests.

As for "cost," there are few Ph.D.s in today's job market who would think that failure to be offered a higher salary than a master's candidate is grounds for rejecting an offer. In short, we have a tragic situation where only 8.5% of community college hirings have the doctorate although every vacancy could be filled with one. Not that it should be - for there are often better M.A.s. But even if only one in three jobs went to Ph.D.s - which might still be a little surprising - that's almost a 400% gain over the present situation. Penalties for overqualification are penalties on the community college system. There are indeed vocational teaching jobs in areas where Ph.D.s could not be found. We are facing a real problem with meeting quality in the California Community Colleges, well put in the title of a September, 1972 article in Change: the Magazine of Higher Learning, "Community Colleges: the Coming Slums of Education?" De facto tenure for poor quality faculty will guarantee slums.

The two major loopholes in the system of using personnel selection to insure curriculum quality, then, are first, that the personnel selected are not as good as they could be - and second, that the best intellectual quality is not a guarantee of good instruction.

The minor loophole is tenure. There is no doubt that tenure protects some incompetence. But systems without tenure protect plenty of incompetence, too. The better the system of making appointments, whether tenured or not, the lower the probability of poor tenure faculty. The tenure system at UC requires about seven years of academic performance before the choice is made - four years more than for high school teachers. That's as long as makes sense. Tenure itself makes sense for one reason above all others and that is to protect unpopular views from censorship by political programs. The college is of limited use to the society's drive for survival unless it can harbor advocates of the major alternative points of view between which the young adult has to choose. If every liberal government can fire every conservative faculty member, and vice

versa, there is soon no one left except the mindless and the treacherous (as happened at the once great University of Vienna in the 1940s). The most fundamental difference between a dictatorship and a republic is the fact that free citizens choose their position - and one cannot choose if one of the alternatives is never presented thoroughly and enthusiastically. The costs of tenure have always been very small in those parts of the system where the selection of staff is made with care; in the rest, the same costs will occur without a tenure system. Insurance is never free, but it is sometimes a bargain.

The one situation where tenure hurts badly is in an ideal one; if we were able to introduce a really good system for reevaluation of all faculty, then we would have good grounds for wishing to replace some who would be protected by tenure. But realistically our best bet is to improve the present selection procedure and make sure that the new blood is good.

DIRECT APPROACHES TO CURRICULUM QUALITY

Although the usual method for getting quality in the content of courses is the indirect one of getting teachers of high intellectual quality, it is neither the only one nor the best. In K-12 education, there have been two decades of curriculum innovation based upon a determination that the existing curriculum was lacking in quality and the adoption of task force approaches to develop better curriculum materials. This approach has not been characteristic of the college level. There are two fairly significant exceptions. One is the professional schools, where California, amongst others, has made some effort to commission informed criticism - for example, the Coordinating Council's early efforts. The other is the undergraduate biology curriculum where CUEBS (The Council for Undergraduate Education in the Biological Sciences), operating under fairly heavy government subsidy, has been engaged both in curriculum production and extensive dissemination activities through a newsletter.

However, one-man or two-man studies of this type are interesting rather than definitive, unless they report on a certain kind of investigation. For this is an area which calls for what is known as a needs-assessment in the evaluation business, a difficult and extensive task. Without going into details, some obvious points are the following:

(a) A serious attempt to state the objectives of the course or curriculum in checkable terms is almost essential before it even gets into the ballpark for evaluation. This is lacking in over 95% of all offerings in California higher education.

(b) Given the objectives one can then look for a rationale for teaching to these goals. The obvious rationale is showing that they fulfill a need.

(c) A need is demonstrated by showing dysfunctional consequences of the lack of capacity to perform these objectives. (A weak sense of 'need' can be shown by showing a demand. But there can be needs which are not recognized and demands which are artificial or mistaken.)

(d) Dysfunctionality can be identified only in terms of a framework of justification for this particular component of education. For example, it's dysfunctional for a physician not to be able to evaluate the statistics used in

drug studies, because there is a justification for training competent medical practitioners.

(e) Although it's easier to see the justification for the practical parts equally compelling justification can be given for the components that educate the citizen for voting in a complex democracy, and the researcher for discovering solutions to extremely difficult problems with significant possibility of social benefits.

(f) The preceding is pretty much common sense though doing it in practice turns out to contain many traps for the beginner. The real expertise comes in when we start asking whether a particular component or emphasis of the course is justified, by comparison with some of the competitive alternatives. To settle this requires considerable subject-matter knowledge as well as knowledge of the competing teaching materials and strategies.

It is amazing what a salutary effect can result from a systematic critical appraisal along these lines of all the subjects taught in the U.S. Probably reading has occupied more teaching hours and research hours than any other subject. Yet - as a 1972 Educational Testing Service study demonstrates - we have essentially no functional literacy tests, no knowledge about what makes a good reading teacher better than a bad

one, and no tests that identify good reading teachers. Additionally, we have no primers whose vocabulary passes the first test of content - namely that it should reflect the speaking vocabulary of the children who use it (in order to avoid adding an unnecessary burden to the decoding task).

The professional schools are little better. The switch to the case-study approach in law schools was a triumph - but it has since become its own problem. The emphasis on memory for minutiae in medical school has been softened but no functional analysis of the knowledge/needs of the practicing physician has ever been done. And so on.

Content, then, is not beyond evaluation and vast improvement. We have talked of courses. But there are breadth requirements and sequencing requirements, etc., too, all of which are handled in the most arbitrary manner. It is hard to avoid wishful thinking as one considers what a lot of red tape and dull facts could be exorcised from the student's life by a serious approach to the evaluation of content. But now

we should turn to the question of who gets the content, such as it is.

QUALITY OF THE DELIVERY SYSTEM

There are two dimensions of the delivery system that require evaluation. They might be called the efficiency dimension, which is concerned with the cost-effectiveness of the procedures for getting higher education to those who receive it, and the service dimension, which is concerned with the question of who gets the delivery. The plain fact about California higher education at this stage is that its delivery system has tended to optimize efficiency considerations at the expense of service considerations. It is a service plus that a serious attempt has been made to arrange for admission, tutoring, and adequate curriculum for some minority group members. But this can scarcely compensate for the fact that a very large number of students who would otherwise attend are prevented from doing so for reasons of cost.*

The delivery problem is of course not independent of the quality of content problem, since no case can be made for delivery of a poor product.

---

* The estimate of 10,000 in 1971-2 has been given for the public four-year systems. This figure will, of course, be greater if the private universities are included.

The discussion of this section therefore proceeds on the assumption that quality considerations have been or are being taken care of in an evaluation system. In particular, it assumes that the major quality obligation - the one that applies to most students, and in substantial part to all students - is to provide the skills, knowledge and perhaps attitudes required for a satisfying life as a contributing member of a democracy.

Broadly conceived, the university delivers not only education but also employment. It is natural to include this aspect of the delivery system under service. It's clear that discrimination in employment has been center stage in recent years, and most of the more obvious steps towards improving the situation have been taken. The only comment one can include here is to strongly endorse the principle of compensatory justice, which implies that affirmative action programs must actually discriminate, to a modest degree, in favor of the minorities who have been previously exploited. It is a necessary consequence of this that other groups will be discriminated against, and if one focuses one's

attention on that fact alone, one is likely to feel that this is an evil. But although it cannot be justified in the long run, one cannot compensate for an advantage that has been illicitly obtained by merely arranging that the rules will be strictly enforced hereafter (since that will preserve the advantage). There must be a period during which the advantage goes the other way, in order to obtain a just situation. This principle has long been recognized in Veterans Preference Laws, and applies no less to the compensation of oppressed minorities and women.

In terms of education, the salient fact is that the same minorities suffer, because they tend to be at the lower end of the economic scale and the major discriminatory activity of California higher education is against those who are not well off. Even though the community colleges, for example, charge little or nothing, the fact remains that textbooks, supplies, transportation, housing and food are simply a debit for the student and must be added to any income that he could make in the job market to provide the total opportunity cost of going to school. That cost

is very high, far more so at the universities, and it can really only be handled by a scale of scholarships or loans that goes far beyond anything presently envisaged. It is not being said the state should do this given the other demands on its services, but it is being said that it should with other things being equal. Given the present situation of underemployment for example, it might well be argued that financial support for attending college should be provided as a small supplement to welfare payments. Perhaps, eventually the increased tax yield resulting from access to jobs with technical prerequisites would provide a net benefit to the state, not just in the very long run, but within a year or two of graduation.

At the moment a kind of political confidence trick is being carried on with respect to the service dimension which increasing education of the public, and perhaps of legislators, should bring to an end. It can be illustrated most graphically in the case of the University of California, but applies elsewhere in California higher education. The political stance is that

"all qualified applicants who are residents of California" will be admitted to U.C. The deception lies in the fact that the applicants will only be guaranteed admission to "some campus," which may be several hundred miles away.* We know that such admissions are frequently turned down. We must not forget that for residents of e.g., Eureka, the nearest U.C. campus is often financially inaccessible because of the cost of living away from home. Only new campuses can serve these needs. Any evaluation of U.C.'s service to the population of California must regard the financial and geographic barriers as making the institution extremely elitist.

It might be thought that this criticism applies much less to the California State University and Colleges and even less to the California Community Colleges. This is true only for the student who is interested in the relatively few areas that are served by all community colleges. Curriculum restrictions are very severe with respect to the U.C.-aimed student, though quite good as vocational training for many jobs. (We need some very careful study of the transfer

* Quotes from Regents' statement of July, 1972.

learning costs.) But the basic problem remains; if the U.C. campus is too far away to attend, why bother with pre-professional work at the community college at all? Inaccessible services are not services at all.

A broad look at any evaluation procedures would certainly require attention to another substantial segment of the population which is currently served rather poorly, namely those who require in-service training or advanced education, collateral with holding a job but not necessarily limited to improving job skills. In short, the so called "adult education" group. Certain dimensions of this group are served spectacularly well by particular community colleges. It is always a pleasure to visit one of these campuses in the evening and see the extent to which its facilities are being gainfully employed by working citizens. But when we examine the situation in the four-year colleges and graduate schools, it turns out to be much less satisfactory, especially with respect to the university system. Here we find that no money from the general tax roll is being provided at all, with the result

that very high fees are charged for extension courses, leading to the same problem of economic (and hence racial, sexual and age) elitism that we have previously discussed. The contrast with other states is nowhere more marked than there. This situation is exacerbated by present restrictions on intermittent attendance at college, which has the effect of another economic barrier.

There is certainly a chain reaction problem with respect to any changes in the system, or in interfaced systems. For example, a decent overhauling of curriculum content and teaching procedures at the K-12 level could result in eliminating a substantial part of the kind of basic political education curriculum that is obviously needed at the college level today. In this sense, California higher education is to some extent a remedial education system for deficiencies in the K-12 system. Correspondingly, adult education is remedial with respect to higher education; not entirely indeed, because its users' needs do change through changes of location and interest, but certainly to some de-

gree. Hence it would seem appropriate to provide substantial funding for adult education as long as it's doing a job which should have been done by the colleges, given that we believe that job is valuable to society. It's possible that eventually the size of the market for adult education, and indeed for higher education might be substantially smaller, given radical reform of the K-12 system. It's also possible that our needs for new technical training, and new dimensions of education for citizenship (the latest example is environmental education), will emerge to make up some of the difference. At the moment, they have to displace other elements in the curriculum, which is not altogether bad given the poor quality of much of it, but this would certainly be bad if curriculum had been reasonably reformed. So in evaluating the present system one must evaluate its capacity for handling substantial changes in the needs of the citizen for higher education. The flexibility it has at the moment is something it would not have if some of the other criticisms were met.

The service dimension of the delivery system, therefore, is not an outstanding feature of U.C.

(in particular, and California higher education in general, though to a lesser extent) despite some explicit attention given to it by President Hitch with his concept of lifelong learning which may lead to reform.

On the other hand, the efficiency dimension of the delivery system has been subject to greater scrutiny, and is a dimension which is much more easily improved in the light of many other examples in the business sector, etc. Nevertheless, no systematic study has been undertaken of California higher education's delivery system as a whole, and once satisfactory standards of service have been set up, one can then go about looking at efficiency with respect to them. One of the major defects of previous reports on the delivery system has been their almost exclusive focus on the question of efficiency, given the present standards of service. Efficiency considerations will have to be reevaluated when and if an improvement in the service standards is made, because one will be dealing with an essentially new system. It should be noted that improved service means the cost will go up per

credit hour delivered, because one is reaching into the areas where more effort has to be made to contact and sustain the students. This must not be read as a decline of efficiency, any more than a company that enters a marginal market should regard increased unit costs per sale as a sign of declining efficiency. It is simply a sign that they have decided to enter a market that is more expensive to service. If that service is worthwhile, then all that can be said about efficiency here is that it costs a little more.

Within the broad framework of service quality discussed so far, there are a number of other questions of importance that would have to be considered in an evaluation of the state's higher education. For example, there is the question of internal justice in the system, by contrast with "external" (that is, the justice in admission appointments we already considered). This includes disciplinary and other arbitration issues for students, staff and faculty. To a large extent the courts act as watchdogs, though the delays in achieving justice often result in

great personal hardship. An excursion into ways of reducing the lag here would take us too far afield from the main task, however.

Again, one should take very seriously the evaluation of the quality of life on the campuses. The sense of anomie does not benefit either learning or loyalty, and it is caused by many small factors, all of them remediable. A typical example is the labelling of buildings. On the older campuses, and on some newer ones, this is done sporadically because "everyone" knows them. The freshman immediately realizes he/she isn't thought of as being anyone, and spends irritated, furious hours wandering aimlessly.

A cure is easily provided for all of these problems. Make every administrator, including each chairman and new appointee, go through registration and a few days of following a class schedule. Get them together to pool their reactions in order to plan change, and make their next promotion or raise dependent upon their success in getting the changes made.

But the evaluator's task is to identify the trouble; he/she does not usually have the power to get it changed.

Still, the quality of life for all members of the campus communities and their neighbors stands clearly in need of improvement and could well be evaluated systematically.

EVALUATION OF RESEARCH

Research has two main functions in higher education; it provides an internal service and an external one. The internal service is to the institution itself and includes three main goals of which we have already considered two. These are:

(a) to ensure the updating of knowledge in each area the institution covers with its teaching program, thereby keeping curricula current and keeping faculty brains in condition to train thinkers in general and researchers in particular.

(b) to assist in recruiting faculty who need research facilities and company. The bet is, again, that it's necessary to have some faculty doing (and interested in) research if any sense of the state of the frontiers of knowledge and the spirit needed to extend them is to come across. (Note that only one in eight students in California higher education are enrolled in the U.C. system, where most of the research is done - hence there is no suggestion that <u>most</u> faculty should be doing this.)

(c) to gain funds, prestige and facilities from external sources. The greater part of the science and math facilities at UC and a good slice at CSUC (have been federally funded because of the contribution to the <u>national</u> research pool. Again, this helps with recruitment.

The external services provide benefits to institutions and individuals outside higher education. Most notable are:

(a) Applied research gains, e.g., to California agribusiness, NASA, etc.

(b) Pure research gains for the fund of knowledge and understanding about the world we live in and ourselves. Remembering that the pure research of today is the savings account on which we draw when the crisis arises tomorrow. And, of course, a very large part of pure research is externally funded.

(c) Personnel gains for the state, industry, and nation in terms of expanding the pool of skilled personnel; and to the individuals themselves whose vocational mobility, and economic wishes are fulfilled. Given the pay off to the state and nation from differential tax gains, and the pay off to industry from lowered on-job-training costs, it seems entirely appropriate that some tax revenue be used for research funding.

The evaluator, apart from identifying the dimensions just listed, can go into investigations of each or all. But it is hard to argue for the urgency of this task. In an earlier version of this paper, there was no <u>separate</u> mention of research at all. Its importance will show up in any adequate evaluation of teaching

and service functions, and the negative public connotations of this dimension, as a luxury system, combined with the record of overemphasis on it in selecting and promoting teachers, makes it a poor argument for the system. The most obvious and probably popular alternative to having California higher education do research would be to contract research that the state or the nation needs to independent research units. It is crucial to understand the connection between research and the other functions of higher education. The way to do that is only to evaluate the other functions. Then, the research will show up to the extent - and only to the extent - that it services the other functions. But this may be somewhat too extreme a remedy.

What about the quality of research? There is considerable plausibility about the hypothesis that it is comparable to that in other systems of higher education in this country. For the horizontal mobility of faculty within the U.S. is so great that a major state can scarcely afford to lower

its standards since it would rapidly become a backwater of inferior talent. That has not, so far, happened in the UC system, though its eminence has certainly diminished, judging from the increase in turn-downs of offers. It is a quite different question whether research is as good as it should (i.e., could reasonably) be, either with respect to the funding it has or with respect to the ideal level of funding.

A related but partially independent question involves the quality of research compared to that in comparable foreign systems, e.g., the Australian one.

To answer such questions with a useful degree of reliability is possible but neither conceptually nor financially trivial. Very little has been done in the way of systematic output evaluation of research. A serious beginning was made in the course of developing a system for evaluating the educational Research and Development centers for the U.S. Office of Education by a team involving Gene Glass (Director of the Laboratory for Educational Research, University of Colorado), Robert Stake (Associate Director, Center for Instructional Research and Curriculum

Evaluation, University of Illinois), Dick Schutz (Director, Southwest Regional Educational Laboratory) and others, under the chairmanship of the present author. But the practical details have never been sorted out because the cost is high. We settled for a rather crude use of peer-group ratings. Now that system is exactly what is now used in California in the course of evaluating faculty, since they are to a large extent evaluated on research (where they are evaluated at all). And we have discussed that extensively in earlier sections.

One might conclude by saying a word about what constitutes "good research." Like "good higher education," it cannot be easily analyzed - but it can be analyzed. The two ultimate grounds of merit for research are social gains (e.g., improved cancer therapy) and knowledge gains (e.g., unified field theory). If the latter had to be justified solely in terms of long-run payoff for the former it would do very well. But it can also be given some independent justification in terms of the desire to know, the curiosity drive.

It is clear that some pure research is an

exercise in futility and expensive to boot. The only problem is deciding which, since there are many notorious examples of the most useless research becoming highly significant. Given that California contributes very little of the cost of pure research and that it pays off in other dimensions than pure knowledge, it seems exceptionally unimportant to work very hard on refining its quality as research. The cost of doing so would likely not be returned as savings. But in general, one must conclude that the UC component overweighs its status compared to the teaching function, and the best way to redress that balance is to correct the evaluation of teaching as suggested earlier.

PREREQUISITES FOR AN EVALUATION SYSTEM

It is clear enough that California higher ecucation is in serious need of an evaluation system. But evaluation systems are not really free initially, and it is pointless to install them just so as to be able to say that they're there. If their results are to be used for maximum social gain, one must first, before installing one, examine carefully the machinery for implementing the results from such a system. And, perhaps more important than any machinery, one must look into the question of motivation.

It is a fundamental truth about objective evaluation of system performance, that everybody thinks it's a great idea for all systems except their own. Evaluation, in short, is for others. This negative attitude arises partly because of the net psychological resistance to criticism of one's own performance, and partly because of the resistance to the effort that is involved in any change. Evaluation is both disruptive and (frequently) critical.

On the other hand, evaluation is a necessity

in order to improve, or even to know one's needs to improve. To somebody with substantial orientation towards maximization of job performance, evaluation is seen as the left hand to go with the right hand of implementation. Externally imposed evaluation almost always creates resistance, and converts what should be a beneficial process into a political infight. It is nearly always essential, therefore, to involve the prospective evaluee - which may be a system - in the process of planning the evaluation system.

In California higher education, there are some very powerful forces that are potential foes of an evaluation system. For example, the unions are likely to oppose any systematic attempts to evaluate job performance, partly because of the long history of abuse of such procedures in the labor-management field, and partly because such systems inevitably uncover substantial shortcomings of performance and thus threaten job security. Administration is no less threatened by critical scrutiny and even students get worried by the thought of someone besides themselves doing the evaluation of their institution.

In the long run it is essential that all these groups have input to any comprehensive evaluation system, or it will not achieve maximum utility. This has led the accreditation commissions to shift towards the role of stimulating self-study. But our task in this paper is only to propose the first steps. Those steps, in order to insure continued support from the Legislature, must involve straightforward attacks on serious problems, with clear answers. The implementation of those answers will require state support - but also continuing evaluation. The continuing system will have to devote considerable effort to reactions of the many components of the higher education community; but to start with that, when no "product" (i.e., recommendations or results) is available would be to risk bogging down in process instead of getting on with production. This is not to say that no efforts at liaison and input should be made by the interim system. On the contrary, good evaluation research absolutely requires it - one can hardly evaluate computer assisted instruction without student input. But this is not political input.

In the long run, then, political input is essential - evaluation is de facto a political instrument - but in the short run it is of secondary importance.

NATURE AND COSTS OF A MINIMAL SUPPLEMENTARY EVALUATION

It will be apparent from the preceding that the task of evaluating California higher education is not being performed adequately. There are a great many ways in which the substantial gaps could be filled in, but it may be useful to provide a few suggestions as to the appearance and cost of an evaluation system. Instead of describing an ideal system, I will here describe a minimal system which would be expected to pay for itself rather quickly - say within three years - and thereafter be expected to make very substantial net savings, part of which should preferably be turned into some modest expansion of the evaluation system in order to generate economy or improvement of education in new areas. It should be remarked again that there may be dimensions of higher education that require investigation or documentation because of legislative or public concern - or indeed academic concern - and these activities of an evaluation system should not be regarded as necessarily productive of net savings.

First, a preliminary note on some of the existing apparatus and capabilities. It should be noticed that one of the U.S. Office of Educations's Research and Development centers of particular interest in this connection was located on the UCLA campus a number of years ago, where it continues to be federally funded. This center happens to be the one - out of the 25 or so that were originally funded - that is concerned with evaluation in the educational domain. And it has a division, funded at the rate of several hundred thousand dollars a year, concerned with the evaluation of higher education. Partly in connection with my decision to accept this contract, and partly because of my general interest in the area, I've had occasion to study the production of this division of the center with some care. Its orientation is simply not recognizably the same as that involved in this report. Having thought at first that they might be better suited - because of their substantial collateral support system - than myself to handle the contract, I became persuaded that what they would do would scarcely be responsive to the urgent need of California higher education and the State Legis-

lature. They are concerned more with getting a subjective response to interview questions about felt needs and satisfactions. Such studies can indeed be suggestive of useful hypotheses, but bear the same relationship to the objective question of the merit of California higher education as would a survey of the patients of faith healers bear to the objective question of the utility of the treatment. The FDA is a supervisory agency charged in part with the evaluation of various medications, and its investigations are not based upon customer satisfaction or dissatisfaction though they may be instigated by such reports. It would be just as inappropriate for the FDA to rely upon the mere testimony of the medical establishment, unsupported by investigative evidence, as it would be for an evaluation of higher education to rest upon professorial opinions.

Evaluation in the medical and educational fields require objective systematic consumer-oriented study of the extent to which the unit under study meets the objectively established needs of possible target populations. A small

part of the determination of needs - and an important part - consists in survey studies of felt needs. But it's neither necessary nor sufficient. For example, we can at the moment establish beyond doubt that 20% of the adult population of the U.S., and probably nearer 30% of the school population, is seriously deficient in functional reading skills. It is unlikely that anything approaching this percentage is aware of this deficiency, or, if aware, willing to identify it as a need. Yet an adequate investigation of "functional reading skills" can establish beyond any possible doubt that these are needs, whether or not they are felt needs. For example, it establishes that road signs conveying urgent information, e.g., about dangers and detours, cannot be read by many people in the time during which they are visible to the driver. To make the opposite point with an example from the medical field, whereas large numbers of people might identify themselves as needing milk - perhaps as a result of the well-known advertising campaign - medical facts do not support this claim (unless the remainining diet is deficient).

So the UCLA data can form part - but only a small part - of the necessary evaluation system.

Apart from the UCLA R & D center, the other chief candidates are the Coordinating Council for Higher Education and the accrediting associations. As we have previously remarked, CCHE is doing input evaluation, with the exception of one or two occasional papers on professional curricula. Again, its data and conclusions are useful contributions to evaluation, but no more satisfactory than an evaluation of the medical care system in this country based solely on considerations of the number and distribution of physicians, without investigation of the patients' welfare. As a result of this deficiency, the Coordinating Council is often reduced to making recommendations on a wholly intuitive basis. Their "minimum class-size" recommendations are made without any knowledge of either the student gains from small classes in certain subjects, or the system costs of cancelling certain offerings on one campus. Equally, of course, the many complaints about their recommendations are with-

out justification. It is a sorry state of affairs when a billion dollar business - or the state that underwrites it - is unwilling to spend a few thousand to find out whether any of its practices are legitimate.

The most interesting of the existing evaluation systems is the Western College Association Accrediting Commission. In many respects, this is an ideal kind of evaluation system - but its very strengths render it incomplete. It is ideal in that it is voluntarily supported by the institutions that it evaluates, thereby rendering the problem of gaining acceptance much more manageable. Most evaluators work very hard to get their subjects to realize that evaluation is the main tool for improvement, and where they come round to the extent of actually funding a permanent installation of this kind (fees are $1200 per annum for large institutions) that point has surely been successfully made. In the old days - and it's still the case for many accreditation organizations - the principal instrument for evaluation was a very complex checklist. Today the emphasis is on generating in-

ternal discussion and self-criticism by the whole educational community being evaluated, and making a holistic judgment of the extent to which the institution fulfills its (possibly very idiosyncratic) goals. Here again the emphasis is on improvement and not just judgment. There are many other excellent features of the process, not least the distillation of many years of experience into the instructions and training procedures.

However, there are drawbacks from the point of view of other consumers of evaluation. The very fact that the evaluations are funded by the institutions themselves renders them suspect, and not without good reason. The members of the visiting committees are usually faculty members from the same region, and hence likely to have either personal or regional loyalties at stake. The criticism of an institution's goals is very limited, in the interests of encouraging pluralism, but with the effect of reducing the amount of evaluation at the most critical point. The selection and training procedures are based on experience, but they have not been critiqued by professional external trainers. For that matter,

there isn't much sign of external evaluation of the system by professional evaluators from other fields. In short, there is an air of coziness about the system which won't sell well with a legislature.* Moreover, it is fundamentally defective with respect to output evaluation for the same reasons given before - they cannot afford and possibly lack motivation to obtain data on learning gains. So we still need a supplementary system, but one which can link up with this one in such a way as to improve effectuation and conserve costs.

Another crucial linkage of any evaluation system would have to be with the office of the Legislative Analyst. And the connection with the Legislature itself requires close study. It is now time to turn to one possible model for such a system, wherein these problems receive attention.

What kind of modest start could be made on an evaluation system that would be useful for the external community? We can conveniently group the discussion under the headings of Mission, Organization, Personnel and Budget.

---

* This complaint is even more serious with respect to the "Bureau of Institutional Research" or Director of Analytical Studies that one finds on a particular campus.

Mission:

There are three considerations governing the selection of mission in these early stages. The first is legislative and public needs. This might involve a direct charge to investigate certain factual questions that concern a new or the old Master Plan for Higher Education because of the need for relatively fast determination of its efficacy. (It might be necessary to create a task force under the evaluation unit itself in some cases.) The second is the need to get moving in directions that will pay off fairly quickly in terms of increased educational merit or effectiveness. The third is the need to establish credibility and a modus vivendi with the academic community itself, whose failure to cooperate would in the long run represent a fatal handicap for the evaluation system. I've previously suggested a very natural first project, of considerable value from each of these points of view, namely the evaluation of the operation of the large introductory courses in California higher education. There are already prototypes or relatives - at Berkeley, Santa Barbara and several CSUC sites -

of what appears to be the optimal method of conducting these courses (the "Keller model"). An immediate need is the objective determination of the relative efficacy and cost of this approach on various campuses by comparison with the traditional approach. Apart from these immediate comparisons, however, there is considerable need to look at the system implications of both alternatives. An excellent example of these is the extent to which the extraordinary inefficiency of the present teaching methods results in multiple duplication of coverage in higher level courses in order to insure a reasonably adequate basic knowledge. It would be an important part of the investigation to show how the immediate advantages of the Keller Model could in fact be accompanied by collateral reductions in this redundancy of succeeding courses.

There are other attractive areas for early evaluations - for example, the "remedial" courses which are essentially teaching skills that everyone agrees should have been acquired in high school. The experience in performance contracting in the K-12 system, is directly relevant here. It

can be summed up by saying that a relatively inexpensive support system, combined with choice of first-rate materials, can out-perform the extremely labor-intensive approach of tutorials or very small classes.

It should be stressed that although both these areas offer substantial possibilities of economy by reduction of staff required for doing a successful teaching task, their attraction is equally great in the direction of straight achievement. That is, the products of the better system now within our grasp could really have the skills that at present only supposedly result from current introductory and remedial courses. There is thus a reasonable chance of obtaining substantial faculty support for the change, on the grounds that the required job is being done better, despite the fact that the unions may well oppose such activities on the grounds that they represent a threat to numbers of jobs. That concern, which certainly has a legitimate dimension, can be met to a considerable extent by implementing other aspects of comprehensive evaluation. That is, if on the one hand we can get actual

teaching efficiency in introductory courses up by a standard deviation or two (a rise of one standard would mean that the present average performance goes up to that previously achieved by the 66th percentile), then on the other hand we can probably identify a substantial group of students (or possible students) who are not being adequately served with respect to vocational or citizenship goals. This gives excellent reasons for converting any economies from staff reductions at one point into expanded services at others.

So the early mission of an evaluation unit should involve simultaneous concern with improved efficiency and improved coverage.

Another important target for early concern is the efficacy of closed-circuit television instruction (CCTV), other audio-visual aids and Computer Assisted Instruction (CAI). All of these involve rather expensive hardware often sold on the promise of substantial economies which have rarely held up in practice. It's not irrelevant to remember that IBM closed its CAI division (partly on the advice of the present

author) after putting a number of years and a great many millions of dollars into it. There may possibly be a use for CCTV with the large introductory courses, and there may be certain very specialized uses for CAI where it can be piggy-backed on an existing computer installation. We shall know a great deal more about this after the present National Science Foundation study of large scale use of CAI is completed.*

In the professional schools, the crucial task for evaluation is to determine the desirability of the present time-base and content-base. Such an enterprise challenges - indeed threatens - such a large part of the teaching staff in the medical and legal centers that it has to be approached with great caution. However, the possible payoff is enormous. Enough has been said to indicate the existence of important mission areas. There are many others, higher and lower in the system.

---

* Alan Post has shown a good instinct for the realities of this situation in his often-quoted speech on higher education in 1971.

Organization:

Implicit in the preceding discussion is a considerable number of judgments about this issue. Put the evaluation system too close to higher education and it becomes a creature of higher education, too far and it becomes completely alien as has the Coordinating Council. The outstanding example of what can be done is the office of the Legislative Analyst, but one may be misled here if he underestimates the contribution of the man himself and thinks that that solution to the organizational problem can safely be generalized. My judgment would be that the best solution is to experiment; I propose one arrangement, but feel very happy about the possibility that after 5 years it might be worth trying another.

The arrangement I favor would be similar to that of the Carnegie Commission on Higher Education which tackled a number of interesting problems about higher education in a rather scattered and inconclusive way. Suppose the Legislature funded an Evaluat[illegible] Center for California Higher Education (ECCHE) with a 5-year commitment, subject to termination for gross negligence, but

not subject to managerial interference before that time was up. That a legislature can do this we see from the office of the Legislative Analyst, and it is worth considering the possibility of tying ECCHE to that office. One way to do this would be to have ECCHE under the control of a 3-person board consisting of the Director of ECCHE, the Legislative Analyst, and a representative from higher educatior. This latter representative would be selected and informed by a 10-person committee with representatives from the various components. It would be expected that the director would frequently meet with that whole Committee, which is his or her implementation arm as well as an important source of information and recommendations. Similarly, there would be meetings with legislative committees, faculty and student organizations. But the prime function of the director would be to get the job done, not public relations. It would be understood that requests which would require putting more than 20% of time into liaison activities may always be legitimately denied.

The hope would still be, however, that the

Legislature could regard ECCHE - if it worked well - as higher education's counterpart of the Legislative Analyst's office. Higher education would have two reasons for working with ECCHE that have not applied to CCHE. First, impeccable professional competence at evaluation, as opposed to coordination. Second, a clear commitment to improvement of instruction rather than reduction of redundancy, etc.

There is some attraction about the idea of an advisory committee for ECCHE, including representatives not only of higher education systems but also from the Western College Accrediting Association, Coordinating Council, and other bodies. But I believe these committees tend to become performers themselves rather than facilitators of performance and would prefer to see it left up to a responsible director to arrange meetings with these organizations as and when necessary. It is obvious that such meetings would be essential in the setting up and early problem-selection process.

At the end of five years, the Legislature would be in an excellent possition to see

whether the expense of ECCHE was justified. The alternative model of setting up a permanent evaluation system immediately, would be harder to get through the Legislature, and would in my view have greater difficulty in attracting staff, for reasons that will appear immediately.

Personnel:

There are two distinct levels at which ECCHE might be funded. To adequately service public requests from the Legislature and segments of higher education in the early years would require what we might call Level I funding (based on experience with the demand level at the Ohio State Evaluation Center). To do the minimal job required to pay off with significant gains, but without responding with advice or additional investigations on request (after the first selection of problems has been made), would require what we might call Level II funding. The effect on personnel is quantitatively obvious, but let's look at the type of person required. There are at least four factors bearing on choice of senior staff:

1. Capacity to design and direct the research.

2. Capacity to attract and retain staff.

3. Capacity to present and relate to "supplier" and "consumer" groups.

3. Independence of mind.

These qualities are not only rather distinct but to some degree in tension. For example, good relations with elements of higher education would be facilitated by part-time appointment within the higher education system - but this would present probable costs with respect to independence.

My inclination is to say that most of the required research should be sub-contracted and monitored by ECCHE staff rather than done in-house. This avoids the inertia and complications of the office building and large numbers of support personnel and office machinery. At a later stage, if ECCHE survived, in-house research may be cost-effective; *now*, it's a poor cost-benefit bet.

Under the sub-contracting option, I'd say that Level I funding would require 5 1/2 full-time positions above the secretarial level; Level II would require 4. In each case, at least one position would be part-time, split with an appointment in higher education (preferably one

at UC and another at CSUC). As between the two options, there would of course be a considerably larger difference in sub-contracted work and costs.

Budget

The minimal system (Level II) should tackle three projects (e.g., the Keller Model for Introductory courses, CCTV, and one curriculum analysis). Using some rules-of-thumb of the educational R & D system, and a lot of care, that would run about $250,000 per year. At least one extra project plus considerable dissemination could be expected within the five-year period.

Level I funding is hard to guess. I estimate 1 1/2 extra senior staff members and funds to handle two extra projects with the limitation that these not run more than two years. That is, nine projects could be covered in the five-year span. This would probably necessitate an additional $125,000 per year.

There is an excellent chance of federal support for some of this work, and other state education agencies would also stand to benefit so much that they might chip in. Proposal-writing

and liaison, with agencies outside California, would cost an extra half-time position, but would be a safe bet for recovery.

I always like to propose super-economy versions of evaluation systems, to give a useful baseline, but we're pretty close to the minimum now. The average federally funded educational R & D center carries three projects and costs $1.8 million annum rent-free. We're talking about projects that are well above their average magnitude and I'm betting that good management can do it on 1/7 of the budget. That makes me a little nervous, but I'd risk it since there's room to drop one project without disaster.

Working from the other end, looking for an even more economical route, suppose one simply used the Legislative Analyst's office to let contracts on the three projects. First, you would lose coordination. Second, you'd lose services. Third, the results would either cost more or be unusably bad, on the evidence we have about targeted research in this field. This is an area where only the best is good enough, and the best can be an economy.

APPENDIX A

METHODOLOGY

The first task was of course to identify resources, which might be useful for (a) already published material, (b) verbal input, (c) critique of the report. The obvious candidates for overall assessments or evaluations of higher education, in one form or another are:

A. The Carnegie Commission on Higher Education.

B. The Higher Education Project at the UCLA R and D Center for Educational Evaluation.

C. The Western Interstate Commission on Higher Education (now the National Commission on Higher Education Management Services).

D. Center for Research and Development in Higher Education, University of California, Berkeley.

E. The Senior Accrediting Commission on Higher Education (Western branch).

F. The Junior Accrediting Commission on Higher Education (Western branch).

G. Administrative staffs at the central offices of the major components of California higher education.

H. California Coordinating Council on Higher Education.

I. Various faculty and educational associations including the AAUP, the NEA, the AFT, etc.

J. Various student organizations.

K. Minutes of meetings of the Board of Regents of the University of California, and the Board of Trustees of the California State University and Colleges, and the Board of Governors of the California Community Colleges.

L. A number of individuals from various independent contracting agencies around the country, or in the federal government components concerned with higher education, whose particular expertise was well known in this connection.

M. Members and staff of the relevant California State Legislature committees.

Given the limitations on committee funds and time and the consultant time, it was obvious from the beginning that no systematic treatment of all these sources, let along the very considerable literature - both conventional and fugitive - would be possible. Some selection from the very beginning was going to be necessary, and a drastic selection at that.

The most recent and up-to-date conventional literature in this area is obviously the immense output of the Carnegie Commission, and the files of the three outstanding journals in the area, namely the Chronicle of Higher Education, Change,

and the AAUP Bulletin. Further relevant literature would be the several thousand items referenced in ERIC (Educational Research Information Clearinghouse), and of course the extensive critical literature (Paul Goodman and others) that has appeared on the subject of tertiary education since the mid-fifties. (Thus, without going back to the traditional literature on this - which includes everything from Cardinal Newman's Idea of the University to Frank Aydelotte's books on the lock step system that led to the introduction of the honors program at Swarthmore, and Ortega's somewhat different Spanish approach - the task was going to be one of radical selection and application, rather than comprehensive survey.)

From another point of view, the same conclusion was apparent; for this area of evaluation is essentially an area where the task of providing advice is not one of scholarly summary but rather of management services. A management consultant will not undertake to provide a history of management consulting as a first chapter in his or her recommendations to the client, but will rather try to provide recommendations and reasons

for them that the client can judge for himself, more or less on the basis of the document itself instead of extensive background reading. Since it was already clear from this consultant's background reading that a particular framework of evaluation, which has emerged in the last decade or so as the most desirable one, was scarcely represented in the literature at all - and since it was clear that this framework could be applied to California higher eeucation with potential significant advantages - the task became a more manageable one if seen as the task of applying this framework to the particular educational system under consideration. There remained some substantial amount of updating, reading and interviewing - principally by telephone - to insure the applicability of this framework to the system and, to straighten about a great many details of the factual kind about its application.

The first task of a good evaluator is to be evaluated; in this case the ideal arrangement would be to get a first draft around to everybody on the main list given above, but at the very least some independent critics on that list

should be identified. Modest honoraria were incorporated in the revised budget in order to make the requests something less than a demand for straight charity. It was early concluded that the experience and background of Bob Hind, Director, Western Region of The Academy for Educational Development, Inc., would be invaluable, a decision which was reinforced by the subsequent discovery that the State Legislature was also involving Dr. Hind in several other aspects of its Joint Committee's work. The search for the second consultant went on during the project, on the grounds that one might well uncover some unexpected source of wisdom in the course of communication with various elements of California higher education and its critics. It became clear, however, that there was no better candidate than Clark Kerr, author of The Uses of the University and chairman of the Carnegie Commission on Higher Education. Dr. Kerr was abroad for the central part of the time of this contract, and it was only possible to contact him in late September. He graciously agreed to undertake this task and constituted the second evaluator of

this evaluator's recommendation.

Even though a selection had to be made from the list of possible resource people with whom it would have been potentially valuable to talk, the selection was made with an eye to picking up recommendations of other valuable resource people that might have been overlooked. This hope was facilitated by requesting all informants to suggest others. An interesting fact soon emerged - namely that many of the informants, whether high placed in California higher education or outside it, were immediately anxious to establish the confidentiality of their communications, or rather were immediately anxious to establish that these would have to be either confidential or extremely restricted. Obviously the value to the State Legislature was greatly increased by getting all the input that was possible, rather than restricting oneself to offical handouts; so it was immediately agreed, whenever the subject was raised, to respect the confidentiality. These "chain-research" procedures, according to which each informant was asked to refer the consultant to other informants, rapidly proved to be turning up very

little in the way of new names or new suggestions that bore on the contract. The reason for this was simply that there is no tradition at all of output evaluation in California higher education, and it was a matter of only a few telephone conversations to identify the rare exceptions to this. Thereafter, people were simply repeating their concern with various measures of input evaluation such as duplication of effort, classroom contact hours, etc. The rest of the task consisted in assembling the materials in a suitable form, and trying to write up a report that would exhibit some continuity of thought and some cumulative impact of argument.

APPENDIX B

ACADEMY FOR EDUCATIONAL DEVELOPMENT, INC
a nonprofit planning organization

September 22, 1972

Professor Michael Scriven
131 Tamalpais Road
Berkeley, California 94708

Dear Michael:

As always, it's a pleasure to work over a text of yours. I'm kept engaged whether I agree or not.

I like the idea you lead up to. In fact your plan, or some variant, seems to be the only way to ride herd on the system constructively. The comments below, then, are really addressed to elements of the paper that don't seem to be leading productively to your conclusion. Here are some quick reactions:

1. We are trying to educate laymen in this exercise. Therefore I believe there is need for some rather basic discussion of evaluation theory, if there is such a thing: the need for clear or agreed goals; the capacity to measure performance against those goals, and unexpected outcomes as well (your "goal-free evaluation" argument); the place of summative and formative evlauation (e. g., feedback to instructors in good student evaluation schemes).

2. There is almost no mention of the role of quantitative data. On the one hand is the need for truly comparable input, throughput, output data (we can't even get an FTE figure now). On the other is the fallacy of the measurement types trying to quantify outputs (look at the mess Jencks is in trying to count everything). I refer specifically to WICHE and ENCHEMS which are nowhere mentioned. You're probably right not to propose any reliance on their creations, but they must be dealt with.

3. There seems to be more criticism of the present workings of the system than necessary. The paper should not be, of itself, an evaluation of what's going on. I have in mind particularly the discussion of faculty load pressure in Chapter 9, and much of Chapter 11 (excluding the discussion of who attends, around p. 54).

4. The whole class hour's discussion (p. 23, f) is important, obviously, but somewhat extraneous to the aims of the paper.

PALO ALTO OFFICE: 770 WELCH ROAD, PALO ALTO, CALIFORNIA 94304 (415) 327-2270

5. You might consider the effect of separating functions in staffing, funding, and evaluation. Specifically, what would happen if we clearly separated teaching and research in this way? This leads to the whole question of rewards, and the fact that people do what they're rewarded for. If you're interested in this line of analysis, I have some stuff on it, summarized in Baldridge (ed) Academic Governance McCutchan, Berkeley 1971 p. 264 f.

6. As a U.C. professor, you need to guard against speaking from the U.S. position -- which I know you don't intend. (See some of the detailed comments below.)

Now for some "one-liners" related to specific parts of the draft.

p. 8 - 6th & 7th lines from the bottom. Actually, the evidence on the relationship of teaching quality and research performance is not at all clear on this point. Some studies show a positive correlation. I'll get citations if you want. A lot depends on the kind of teaching and the goals of the student, in my opinion.

p. 9-10 - Does primary reliance on texts make much sense at community colleges, which represent more than half of CHE?

p. 10 - Suggest you define instructor so the reader won't confuse it with Instructor.

p. 10 - A couple of sentences describing Keller's approach would be useful.

p. 19 - lines 9-10. You're a bit harsh on student evaluation, since there are some useful systems -- Washington and Purdue were the best examples a few years ago. At the very least, a well designed and administered student evaluation provides useful feedback -- formative evaluation.

p. 23 - The whole class hour bruhaha is so loaded that I doubt the value of giving the topic this much attention. I hope there aren't many people around who think that keeping score on class hours is evaluation.

p. 25 - lines 1 & 2. I think operations research should be dropped from the list. It doesn't seem to fit with the others in character or chronology.

p. 29-31, & Chapter 6 in general. Unionization is not mentioned here (only on p. 60, to my recollection). When it happens it will surely bring organizational and behavioral changes.

p. 32-33. Outside grading is, I agree, a very effective device, but has it ever been used in a mass system? Can it be? I'm not saying it can't, but it may have to be modified (mechanized?) before it can be useful with 800,000 enrollment.

p. 35, line 7. Is the community college ("the great span of CHE") instructor really "left on his own"? Also, the last paragraph on p. 35 rather contradicts this assertion.

p. 54, lines 15 & 16. I believe the practice of requiring extension to pay for itself is rather widespread, (even though highly unwise and unfair).

p. 62 f. This tells me less about what actually is being done than what is not being done. Reference to student evaluation, reviews by State agencies, would be a minimum because I doubt that you should take your time to survey the situation in depth.

p. 73-74. I don't like the overhead argument because it is so hypothetical (universities don't in fact try to make money as a central goal). Furthermore, I'm pretty well convinced that universities do not profit from overhead charges. They just spend it foolishly on a bureaucracy that has been created to deal with another bureaucracy's absurdities.

Hope this helps. I'm on call if I've not been clear, or if more comment would be useful.

Best regards to you and Mary Ann.

Sincerely,

Robert R. Hind
Director, Western Region

RRH/mm

P. S. I'm still one behind on your other project. Let me know if it's too late to comment.

CARNEGIE COMMISSION ON HIGHER EDUCATION

Clark Kerr, Chairman

Ie Ashby
Ralph M. Besse
Joseph P. Cosand*
William Friday
Patricia Roberts Harris
David D. Henry
Theodore M. Hesburgh, C.S.C.
Stanley J. Heywood
Carl Kaysen
Kenneth Keniston
Katharine E. McBride
James A. Perkins
Clifton W. Phalen
Nathan M. Pusey
David Riesman
William W. Scranton
Norton Simon
Kenneth S. Tollett

1947 Center Street
Berkeley, California 94704
(415) 849-4474

September 29, 1972

Professor Michael Scriven
131 Tamalpais Road
Berkeley, Ca. 94708

Dear Michael:

It all looks fine to me But suggest attention to (1) "student satisfaction" (they are very intelligent consumers!) like in the Trow study (the research university shows up very well); (2) data on % getting merit increases and going on to tenure as in the recent HEP study--good indication of quality review; and (3) inter-state cost data for somewhat comparable systems--California looks good!--see our report on "Capitol and Campus" for type of data available on state expenditures.

Best regards,

Clark

Clark Kerr

*On leave from March 1, 1972